ONLINE APPENDIX Ideological Moderation and Success in U.S. Elections, 2020-2022

Michael A. Bailey *Georgetown University*

Benjamin F. Reese *Georgetown University*

Appendix 1: Measuring ideology from text

Model The statistical model builds on Slapin and Proksch (2008)'s Wordfish model (see also, e.g., Lauderdale and Herzog (2016) and Temporao et al. (2018)). In the Wordfish model, the number of times candidate c uses term w is governed by a Poisson process in which

$$Pr(Y_{cw} = k) = \frac{\lambda_{cw}^k e^{-\lambda_{cw}}}{k!} \tag{1}$$

where $\lambda_{cw} = e^{Z_{cw}}$ is the expected value of the counts (and the variance, as well) and

$$Z_{cw} = \beta_{0w} + \alpha_c + \beta_{1w}\theta_c \tag{2}$$

The β_{0w} parameters reflect the propensity of term w to be used across all candidates and the α_c parameters reflect the propensity of candidate c to use all terms (i.e., the verbosity of the candidate). The β_{1w} parameter reflects the extent to which candidate ideology (θ_c) is associated with the word's usage. A positive β_{1w} indicates that conservatives use the term more often; a negative β_{1w} indicates liberals use the term more often.

We estimate separate models for the website and Twitter data. In the analysis, we use the weighted average of the Twitter and website ideology where the weights are based on the number of words the candidate used on web and Twitter, normalized to differences in overall word usage in our data set on the two platforms.

Word usage varies across parties in ways that undermine a direct application of Wordfish or related algorithms. The figures below illustrate the problem by showing the raw data on word counts plotted against the roll-call based ideology for two sets of terms. Figure A1 shows words favored by more extreme candidates in both parties. Use of the term "MAGA" is plotted in the upper left panel. Clearly, extreme candidates use this word more often than moderates. If we use the standard Wordfish model, we will estimate the relationship between the use of the word and ideology characterized by the gray line: a weak positive relationship indicating that conservatives are a bit more likely to use the term.

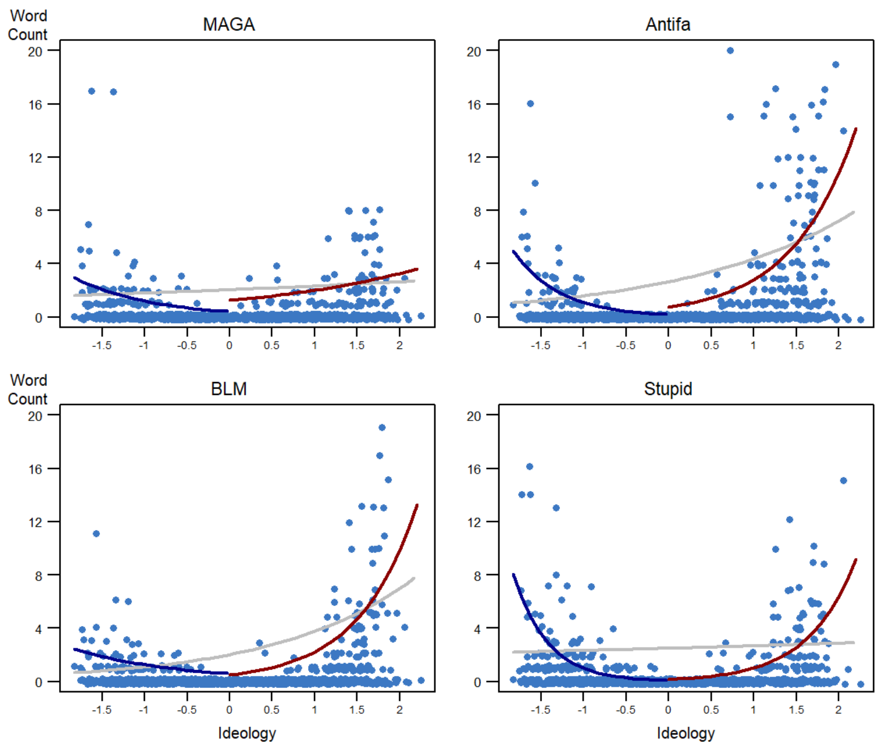

Figure A1: Extreme words

To illustrate the potential for improved fit, we estimate separate count models for Democratic and Republican incumbents using their roll-call based ideology and plot fitted values with blue and red lines, respectively. For Democrats, use of MAGA is associated with liberalism while for conservatives use of the term is associated with conservatism. The other panels show similar examples for terms such as “antifa”, “BLM” and “stupid”.

Figure A2 illustrates another way in which term use varies by party by showing words favored by moderate candidates in both parties. For example, the upper left panel shows use of the word “bipartisan” as a function of ideology. The standard Wordfish model would estimate the weak positive relationship indicated by the grey line, suggesting that candidates who mention the term are more likely to be more conservative. The blue and red lines allow the pattern to differ by party

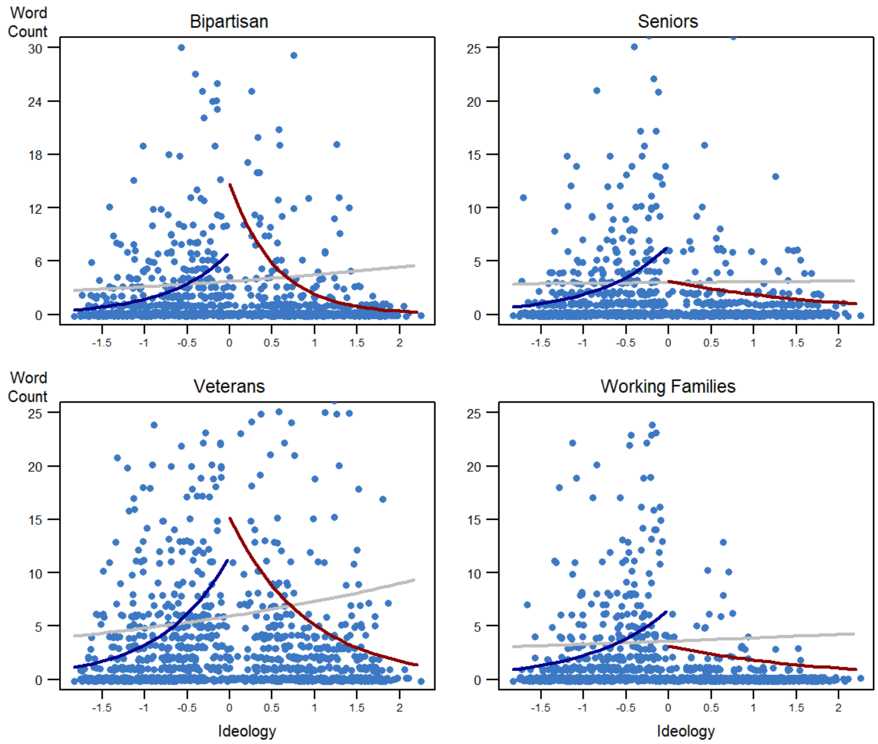

Figure A2: Moderate words

and we see that Democrats who use the term are indeed more conservative but Republicans who use the term are less conservative. Terms such as "seniors", "veterans" and "working families" reveal similar patterns, patterns that are also common in the data.

Bailey (2023) addresses the issue by implementing a filtering algorithm that first identifies whether a term will likely have heterogeneous effects across parties. If it does, term coefficients are estimated separately by party; if it does not, term coefficients are estimated in a single model for both parties. Additional details are available in Bailey (2023).

An issue that arises in text-based scaling models is a proliferation of terms. For example, the 2020 Twitter data has 618,680 unigrams and 27,762,023 n-grams of size four. The website data has 40,846 unigrams and 1,750,258 n-grams of size four. The possibility for spurious associations rises with the dimensionality of the text data, potentially threatening the requirement in text-to-ideology models that "ideology dominates the language used in the text" (Grimmer and Stewart, 2013, 269).

The raw text data is therefore filtered in several steps. First, the data is limited to terms with sufficient term frequency (50 for Twitter n-grams, 200 for Twitter unigrams and 20 for the websites) and document frequency (10 for both sources). N-grams are limited to those identified as collocations by the quanteda package (Benoit et al., 2018). In addition, self-references by candidates and state and regional terms are removed. We retain only terms and n-grams that predict ideology or party among in incumbents in a pre-screening stage. After these filters, the 2020 Twitter data has 16,496 unigrams and 18,752, 2,174 and 199 n-grams of size two, three and four, respectively. The 2020 website data has 4,314 unigrams and 1,609, 137 and 8 n-grams of size two, three and four, respectively.

The estimates used in the paper here are based on a slightly evolved version of the models reported in Bailey (2023). In the original paper, term and ideology parameters were simultaneously estimated for all candidates. The estimates had good properties as described in the paper, but produced differently shaped distributions for Democratic incumbents when comparing Twitter and website data. Additional analysis indicated that this was because non-incumbents tended to use a subset of words (sometimes obscure or uncivil terms) that tended to push them to extremes relative to incumbents. In the current version, the estimates are produced in two steps. First, the term parameters are estimated on incumbent data. Second, the ideology estimates are fitted ideology given these parameters for all candidates, incumbents and non-incumbents alike.

Results For the website data, there are 877 and 764 Democratic candidates in 2020 and 2022 and 778 and 1,050 Republican candidates in 2020 and 2022. We use the website as it was on Election Day or the latest version we scraped before that day. For Twitter, there are 893 and 778 Democratic candidates and 801 and 871 Republican candidates in 2020 and 2022.

Figure A3 shows the densities of the ideology estimates for various groupings and data sources for 2022; all analogous figures look similar for 2020. The upper left panel shows all House members. The ideologies densities for both data sources overlap almost completely. The identifying assumption is that the incumbent ideologies are mean zero with variance 1, so it is not surprising (but not inevitable) that the densities overlap as they do. The upper right panel shows ideologies for non-incumbents. Because the identifying assumption applied only to incumbents it is possible for the incumbent and non-incumbent distributions to vary. The densities largely overlap; the Twitter densities tend to be slightly more extreme. The bottom panels show the densities for Senate and gubernatorial candidates.

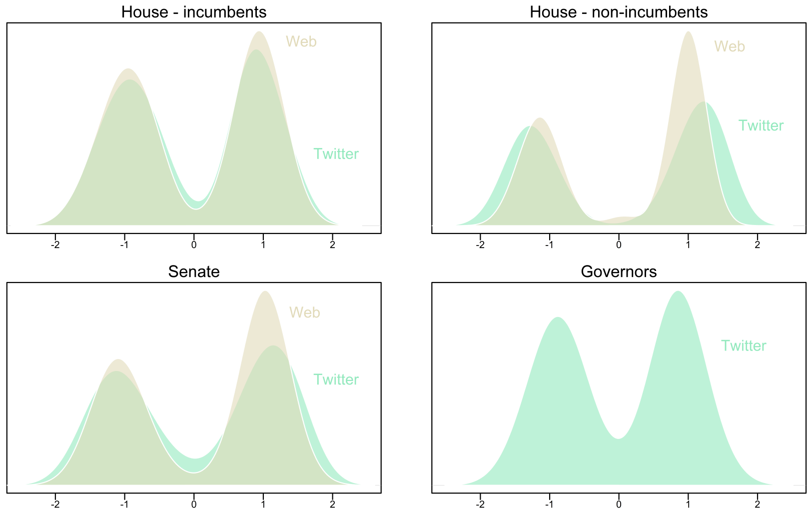

Figure A3: Ideology densities, 2022

Validity To provide a sense of the results from the model, Table A1 shows inter-party terms with the largest ideological valence in the website and Twitter data, respectively. For these terms, a single model applies to Democrats and Republicans. The column on the left shows words that have some of the most negative values of $\hat{\beta}_{1w}$; the more a candidate uses these terms, the more liberal their estimated ideal point. The column on the right shows terms that have some of the most positive values of $\hat{\beta}_{1w}$. These terms comport well with language associated with liberals and conservatives.

Table A1: Ideological inter-party terms, 2020

Web		Twitter	
Liberal	Conservative	Liberal	Conservative
civilrights	alternatives	disparity	balance_budget
color	carry	economic_justice	balanced_budget
communities_color	free_market	eviction	clapper
conditions	freedom	hash_nowarwithiran	hash_antifa
detention	god	healthcare_housing	leftist
disparities	illegal_immigration	immigrant_communities	leftists
immigrant	larger	indigenous	sars
inequality	pledge	labor_rights	secure_borders
police	principles	racial_economic	swamp

right_vote	repealing	tamir	trillion_debt

Table A2 displays the most ideologically informative party specific terms from the website-based models. These terms were identified as having differential party effects in a pre-screening process; the terms in the table are terms with large absolute values of the parameter that connects term usage to ideology. On the left are Democratic terms. The Liberal Democrat terms have negative coefficients, indicating that liberal Democrats use them more than moderate Democrats. The Moderate Democrat terms have positive coefficients, indicating that moderate Democrats use them more than liberals. On the right are Republican terms. The Moderate Republican terms have negative coefficients, indicating that moderate Republicans use them more than conservatives. The Conservative Republican terms have positive coefficients, indicating that conservatives use them more than moderate Republicans.

Table A2: Ideological website terms (truncated), by group

Liberal Democrat	Moderate Democrat	Moderate Republican	Conservative
cannabis	failing	bipartisan	citizen
incarcerated	farming	clean	dr
indigenous	fewer	manufacturing	form
legalization	folks	mental_health	free_market
new_deal	listen	native	freedom
public_housing	lobbyists	police	god
rent	navy	region	pledge
single_payer	pledge	senate	religion
trans	rural	woman	religious
war_drugs	rural_communities	women	required
wealth	secondamendment	youth	wall

Table A3 displays the most ideologically informative party specific terms from the Twitter-based models. The interpretation of the columns is the same as for Table A2.

Table A3: Ideological Twitter terms (truncated), by group

Liberal Democrat	Moderate Democrat	Moderate Republican	Conservative Republican
cancel_student_debt	family_farmers	abt	clapper

fundraisers	family_farms	bail_reform	crap
hash_abolishice	helped_pass	diabetes	drunk
hash_cancelstudentd	ibew	federal_level	endless_wars
hash_dem	lit_drop	funding_help	illegal_alien
hash_medicareforall	navy_veteran	great_american_outd	kinda
hash_notmeus	problem_solvers	miners	leftists
housing_human_right	salt_deduction	storage	opt
jobs_guarantee	tier	tribal	rand
public_college	upstate	tribes	troll
sis	working_across_aisl	village	yall

Table A4 provides results for selected candidates. Prominent conservatives such as Representatives Gosar, Gaetz and Gohmert are on the conservative end of the party. The average ideology for incumbent Republicans is 1.04. Prominent moderate Republicans such as Representatives Kinzinger and Gonzales are left of the Republican average. For the Democrats, high profile moderates such as Representatives Spanberger, Cuellar and Lamb are to the right of the party average. Prominent liberals such as Representatives Ocasio-Cortez and Omar anchor the left. Generally, Republican moderates generally represent districts where Trump had 55 percent or less of the vote while Democratic moderates represent districts where Trump generally had more than 45 percent of the vote in 2016.

Table A4: Selected Candidates

Candidate	District	Ideal Point
Jim Jordan (R)	OH-04	1.91
Paul Gosar (R)	AZ-04	1.75
Matt Gaetz (R)	FL-01	1.71
Lauren Boebert (R)	CO-03	1.63
Marjorie Taylor Greene (R)	GA-14	1.62
Louie Gohmert (R)	TX-01	1.58
Adam Kinzinger (R)	IL-16	0.96
Young Kim (R)	CA-39	0.86
Anthony Gonzalez (R)	OH-16	0.75
John Katko (R)	NY-24	0.43
Brian Fitzpatrick (R)	PA-01	0.38
Anthony Brindisi (D)	NY-22	-0.01
Kyrsten Sinema (D)	AZ-Sen	-0.03
Xochitl Torres Small (D)	NM-02	-0.11

Elaine Luria (D)	VA-02	-0.13
Abigail Spanberger (D)	VA-07	-0.26
Conor Lamb (D)	PA-17	-0.27
Rita Hart (D)	IA-02	-0.31
Henry Cuellar (D)	TX-28	-0.39
Steny Hoyer (D)	MD-05	-0.80
Raphael Warnock (D)	GA-Sen	-0.99
Jamaal Bowman (D)	NY-16	-1.39
Cori Bush (D)	MO-01	-1.41
Alexandria Ocasio-Cortez (D)	NY-14	-1.59
Ilhan Omar (D)	MN-05	-1.62

Table A4 includes several prominent non-incumbents. These candidates had no roll-call record to provide information about them; their language, however, placed them in ideological space consistent with their reputations (and later voting records). Republican candidates Boebert, Taylor Greene and Cawthorn were high profile conservative newcomers to congressional politics. Candidate Kim flipped a California seat for the Republicans, running as a moderate. Candidate Hart is a moderate Democrat who narrowly lost her Iowa race. Candidate Warnock won in a heavily contested Georgia Senate race, using language that placed him squarely in the middle of the Democrats. Candidates Bowman and Bush defeated incumbent Democrats in the primaries and positioned themselves in the progressive wing of the party.

Appendix 2: Statistical results for Figures 4 and 5

We show in the theoretical model in Section 2 how looking at differential performance by Democrats in two races controls for μ_d, which is the predisposition of the district to vote for Democratic candidates. As is standard in fixed effect models, this factor will control any demographic characteristic that is fixed within district across the two races. It is possible, however that demographic features affect the two races differently. For example suppose that

X_d is a demographic characteristic that had β_X^{House} effect on the vote share of the Democratic House candidate and β_X^{Senate} effect on the vote share of the Democratic Senate candidate.

$$\begin{aligned} DemPct_d^{House} &= \mu_d + \beta_X^{House} X_d + \beta \kappa_d^{House} \\ DemPct_d^{Senate} &= \mu_d + \beta_X^{Senate} X_d + \beta \kappa_d^{Senate} \end{aligned}$$

In this case, the difference in performance by Democrats in House and Senate races is

$$\begin{aligned} DemPct_d^{House} - DemPct_d^{Senate} &= (\beta_X^{House} - \beta_X^{Senate}) X_d + \beta(\kappa_d^{House} - \kappa_d^{Senate}) \quad (3) \\ &= \tilde{\beta}_X X_d + \beta(\kappa_d^{House} - \kappa_d^{Senate}) \quad (4) \end{aligned}$$

where $\tilde{\beta}_X$ reflects the differential effect of X_d on House and Senate races. We do not have specific expectations with regard to why a given demographic characteristic may help or hurt Democratic candidates in one type of race (e.g., House) versus another (e.g., Senate), but we include these controls in order to account for additional non-ideological factors that could effect outcomes and potentially be correlated with ideological cutpoint differences.

Tables A5 through A7 present congressional district level results. Columns (1) through (3) show results using all available data. Columns (4) through (6) show results using data limited to House districts in which Donald Trump received between 40 and 60 percent of the two-party presidential vote in 2016.

Figure 5 in the paper presents results from columns (3) and (6), which include controls for incumbency and district demographics. Models with fewer control variables generally produce larger coefficient estimates for the ideological cutpoint difference variable.

The expected coefficient on the Democratic House incumbent variable is positive in these

Table A5: House Democratic margin relative to Biden margin, 2020

Dependent Variable:	Difference in Democratic percent					
	Contested			Competitive		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Cutpoint difference	5.0^{***} (0.74)	3.7^{***} (0.67)	2.7^{***} (0.66)	9.0^{***} (0.97)	5.2^{***} (1.1)	4.7^{***} (0.96)
Dem. House inc.		0.48 (0.40)	1.2^{***} (0.40)		0.92^{*} (0.55)	1.5^{***} (0.51)
Rep. House inc.		-2.1^{***} (0.42)	-2.5^{***} (0.40)		-1.9^{***} (0.54)	-2.1^{***} (0.50)
Black pct.			-0.02^{*} (0.01)			0.02 (0.02)
Hispanic pct.			-0.02^{**} (0.009)			-0.02 (0.02)
Median income			-1.7 (1.3)			-1.2 (1.8)
College pct.			-0.07^{***} (0.02)			-0.10^{***} (0.03)
Constant	-1.3^{***} (0.15)	-0.65^{*} (0.37)	3.0^{***} (0.68)	-2.3^{***} (0.21)	-1.4^{***} (0.48)	2.4^{***} (0.87)
Fit statistics						
Observations	380	380	380	173	173	173
R^2	0.11	0.29	0.38	0.33	0.46	0.57

IID standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*
Note: Median income in $100,000

Table A6: House Democratic margin relative to Biden 2020 margin, 2022

Dependent Variable:	Difference in Democratic percent					
	Contested			Competitive		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Cutpoint difference	6.8^{***} (0.87)	5.4^{***} (0.91)	4.3^{***} (0.85)	11.6^{***} (1.3)	9.8^{***} (1.6)	8.2^{***} (1.4)
Dem. House inc.		0.93^{**} (0.38)	1.8^{***} (0.36)		1.3^{**} (0.61)	2.0^{***} (0.56)
Rep. House inc.		-0.50 (0.37)	-1.0^{***} (0.35)		0.02 (0.53)	-0.33 (0.48)
Black pct.			-0.06^{***} (0.01)			-0.08^{***} (0.02)
Hispanic pct.			-0.06^{***} (0.008)			-0.07^{***} (0.01)
Median income			-5.9^{***} (1.1)			-7.3^{***} (1.7)
College pct.			0.04^{*} (0.02)			0.06^{*} (0.04)
Constant	-1.7^{***} (0.18)	-2.1^{***} (0.33)	2.3^{***} (0.71)	-1.6^{***} (0.25)	-2.1^{***} (0.43)	2.7^{**} (1.1)
Fit statistics						
Observations	378	378	378	166	166	166
R^2	0.14	0.18	0.34	0.34	0.36	0.51

IID standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*
Note: Median income in $100,000

Table A7: House Democratic margin relative to Senate Democratic margin, 2020

Dependent Variable:	Difference in Democratic percent					
	Contested			Competitive		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Cutpoint difference	8.2^{***} (1.4)	4.2^{***} (1.5)	4.2^{**} (1.7)	11.1^{***} (2.1)	6.0^{**} (2.7)	5.7^{**} (2.8)
Dem. House inc.		3.2^{***} (0.94)	2.9^{**} (1.2)		2.7 (1.7)	1.6 (2.1)
Dem. Sen. inc.		-4.4^{***} (1.2)	-4.8^{***} (1.4)		-2.5 (2.8)	-4.3 (3.1)
Rep. House inc.		-1.2 (0.92)	-1.1 (0.94)		-0.83 (1.5)	-0.76 (1.6)
Rep. Sen. inc.		-0.89 (1.2)	-1.2 (1.2)		0.79 (2.8)	0.01 (3.0)
Trump 2016 pct.			-0.86 (4.3)			-17.3 (16.6)
Black pct.			0.01 (0.03)			0.009 (0.06)
Hispanic pct.			0.03 (0.03)			0.02 (0.07)
Median income			6.5 (4.5)			13.2^{*} (7.0)
College pct.			-0.08 (0.08)			-0.22^{*} (0.13)
Constant	-2.0^{***} (0.36)	-0.48 (1.3)	-1.6 (3.8)	-2.8^{***} (0.55)	-2.3 (2.8)	7.1 (11.3)
Fit statistics						
Observations	183	183	183	83	83	83
R^2	0.15	0.31	0.33	0.26	0.34	0.39

IID standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*
Note: Median income in $100,000

models because – all else equal – we expect Democratic incumbents to run further ahead (or less behind) President Biden or their Senate copartisans than Democratic non-incumbents. The expected coefficient on Republican House incumbents is negative for similar, but opposite, reasons. In Table A7, which also includes Senate incumbency controls, the expected coefficient on Democratic Senate incumbency is negative, as we would expect – all else equal – that a Democratic House member will run less far ahead (or further behind) a Senate Democratic incumbent. The expected coefficient on Senate Republican incumbency is positive based on analogous reasoning.

Appendix 3: Statistical results for Figures 6 and 7

In Table A8, the expected coefficients on Democratic and Republican incumbency are positive and negative, respectively, for similar reasons described above. In Table A9, the expected coefficient on Democratic gubernatorial incumbency is negative because all else equal a Democratic Senate candidate is expected to run less far ahead (or further behind) a Democratic governor running for re-election. The expected coefficient on Republican gubernatorial incumbency is positive for analogous reasons. In Table A10, the expected coefficient on Democratic gubernatorial incumbency is negative because Biden should run less far ahead (or further behind) an incumbent Democratic governor. The expected coefficient on Republican gubernatorial incumbency is positive because Biden should run further ahead (or or less far behind) of a Democratic gubernatorial candidate who faces an incumbent Republican governor.

Table A8: Senate Democratic margin relative to Biden 2020 margin, 2020 & 2022

Dependent Variable:	Difference in Democratic percent					
	County level			State level		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Cutpoint difference	5.4*** (1.9)	4.8*** (1.7)	3.9** (1.7)	5.9*** (1.4)	4.9** (1.9)	5.5** (2.0)
Dem. Sen. inc.		-1.3 (1.1)	-0.40 (1.1)		-1.5* (0.88)	-1.2 (1.0)
Rep. Sen. inc.		-2.1 (1.4)	-1.7 (1.3)		-2.1** (0.95)	-1.3 (0.91)
2020		0.39 (0.77)	0.54 (0.77)		0.71 (0.81)	0.33 (0.83)
Black pct.			-0.02 (0.01)			-0.02 (0.04)
Hispanic pct.			-0.05*** (0.01)			-0.10*** (0.03)
Median income			-2.1 (1.3)			13.3 (8.5)
College pct.			-0.03 (0.02)			-0.19 (0.13)
Constant	-0.27 (0.29)	0.87 (0.93)	3.4*** (1.1)	-0.40 (0.34)	0.63 (0.89)	0.13 (2.9)
Fit statistics						
Observations	3,952	3,952	3,948	58	58	58
R^2	0.16	0.21	0.34	0.23	0.30	0.46

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: Median income in $100,000. Weighted by number of votes. Standard errors for county results clustered at state level.

Table A9: Senate Democratic margin relative to gubernatorial Democratic margin, 2020 & 2022

Dependent Variable:	Difference in Democratic percent					
	County level			State level		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Cutpoint difference	18.2*** (3.3)	13.6*** (4.3)	13.6*** (4.1)	18.2*** (4.3)	13.5* (7.0)	15.6** (6.2)
Dem. Sen. inc.		0.78 (2.4)	1.8 (1.8)		0.81 (3.1)	4.8* (2.6)
Rep. Sen. inc.		-3.4* (1.9)	-2.2 (1.6)		-3.4 (3.3)	0.97 (3.0)
Dem. Gov. inc.		2.4 (1.9)	1.6 (1.6)		2.3 (3.3)	-0.70 (2.7)
Rep. Gov. inc.		5.9*** (2.1)	6.5*** (1.9)		5.9 (3.7)	7.0** (3.1)
2020		4.8* (2.7)	5.2** (2.1)		4.8 (3.3)	4.7* (2.4)
Trump 2020 pct.			-16.4** (6.6)			-0.97*** (0.26)
Black pct.			-0.15** (0.07)			-0.42*** (0.13)
Hispanic pct.			-0.10*** (0.03)			-0.33*** (0.10)
Median income			-0.99 (3.7)			-65.9** (26.5)
College pct.			-0.12* (0.06)			0.07 (0.33)
Constant	-0.03 (0.76)	-2.8* (1.4)	12.4** (5.5)	-0.01 (0.96)	-2.8 (2.7)	90.6*** (27.8)
Fit statistics						
Observations	1,427	1,427	1,426	24	24	24
R^2	0.42	0.56	0.63	0.45	0.59	0.86

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: Median income in $100,000. Weighted by number of votes. Standard errors for county results clustered at state level.

Table A10: Biden 2020 margin relative to gubernatorial Democratic margin, 2020 & 2022

Dependent Variable:	Difference in Democratic percent					
	County level			State level		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Cutpoint difference	11.0*** (2.3)	16.0*** (3.2)	14.6*** (3.7)	11.3*** (3.3)	15.6*** (4.6)	14.0*** (4.8)
Dem. Gov. inc.		3.7 (3.1)	3.4 (2.8)		3.7* (2.2)	3.5 (2.1)
Rep. Gov. inc.		3.7 (3.3)	4.4 (3.1)		3.7 (2.4)	6.5** (2.6)
2020		5.2** (2.0)	5.3*** (1.9)		4.6* (2.5)	4.8* (2.5)
Black pct.			0.0005 (0.04)			-0.04 (0.10)
Hispanic pct.			0.04 (0.03)			0.03 (0.06)
Median income			4.9 (3.4)			2.7 (19.1)
College pct.			0.006 (0.04)			0.48 (0.37)
Constant	-0.38 (0.97)	-5.7* (2.8)	-9.6** (3.8)	-0.13 (1.1)	-5.2** (2.2)	-22.9*** (7.2)
Fit statistics						
Observations	2,438	2,438	2,435	39	39	39
R^2	0.22	0.38	0.41	0.24	0.39	0.52

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: Median income in $100,000. Weighted by number of votes. Standard errors for county results clustered at state level.

References

Bailey, Michael A. 2023. "Measuring Candidate Ideology from Congressional Tweets and Websites." Manuscript, Georgetown University and SSRN.
URL: *Available at SSRN: https://ssrn.com/abstract=4350550*

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller and Akitaka Matsuo. 2018. "quanteda: An R package for the quantitative analysis of textual data." *Journal of Open Source Software* 3(30):774.

Grimmer, Justin and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3):267–297.

Lauderdale, Benjamin E. and Alexander Herzog. 2016. "Measuring political positions from legislative speech." *Political Analysis* 24(3):374–394.

Slapin, Jonathan B. and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52(3):705–722.

Temporao, Mickael, Corentin Vande Kerckhove, Clifton van der Linden, Yannick Dufresne and Julien M. Hendrickx. 2018. "Ideological Scaling of Social Media Users: A Dynamic Lexicon Approach." *Political Analysis* 26(4):457–473.